Testing specific hypotheses on the distribution underlying the item scores in the Graded Response Model

Dylan Molenaar, Conor Dolan & Paul de Boeck

University of Amsterdam

IRT workshop, 12th of October 2010, Enschede.

In short:

- Derive the GRM by assuming categorization of a continuous variable (Wirth & Edwards, 2007).
- Discuss the effects of asymmetry in the distribution of this variable (Samejima, 1997; 2000; 2008).
- Propose specific tests for departures from symmetry.
- Simulation and application.

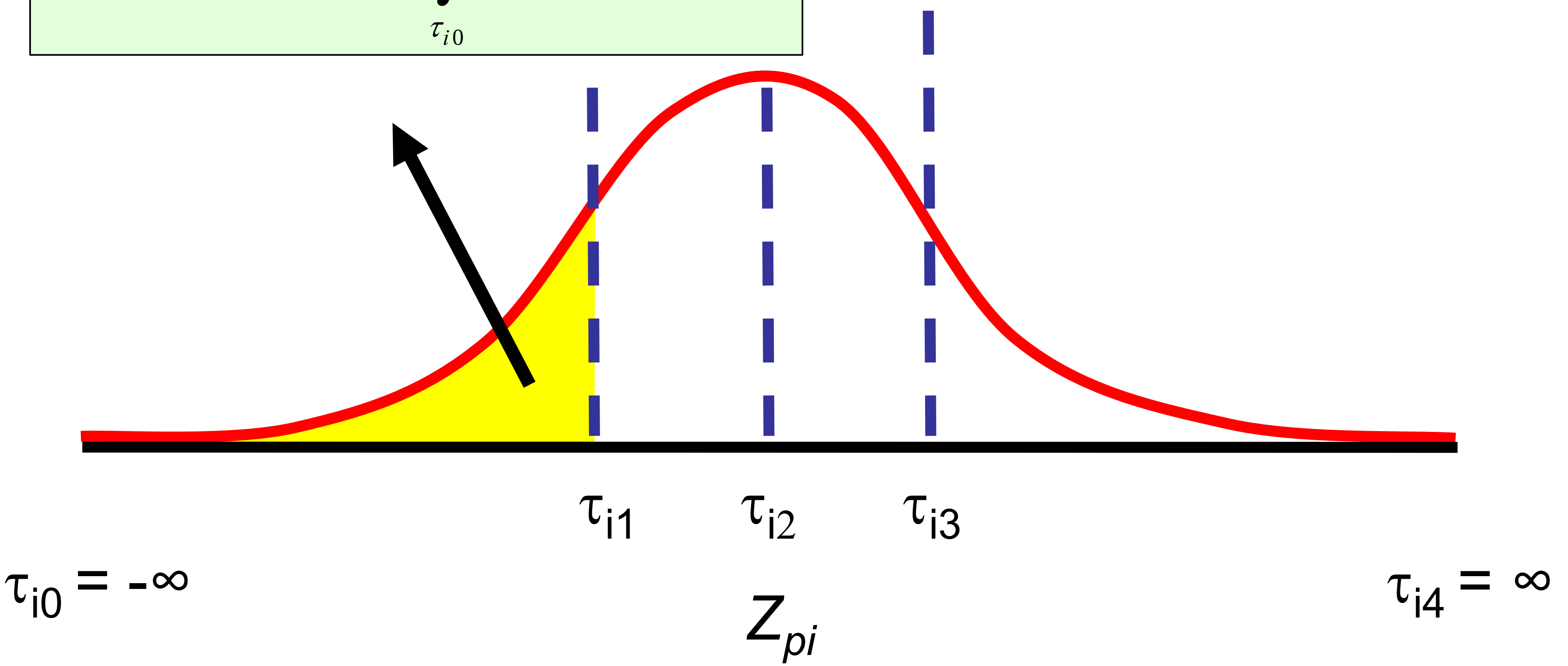
A Derivation of the GRM

(Wirth & Edwards, 2007; Takane & de Leeuw, 1989)

- … the ordinal score of person *p* on item *i*.
- … at specific thresholds.

Subjects score '0'

$$P(X_{pi} = 0 \mid Z_{pi}) = \int_{\tau_{i0}}^{\tau_{i1}} h(Z_{pi}; \boldsymbol{\gamma}_{pi})\, dZ_{pi}$$



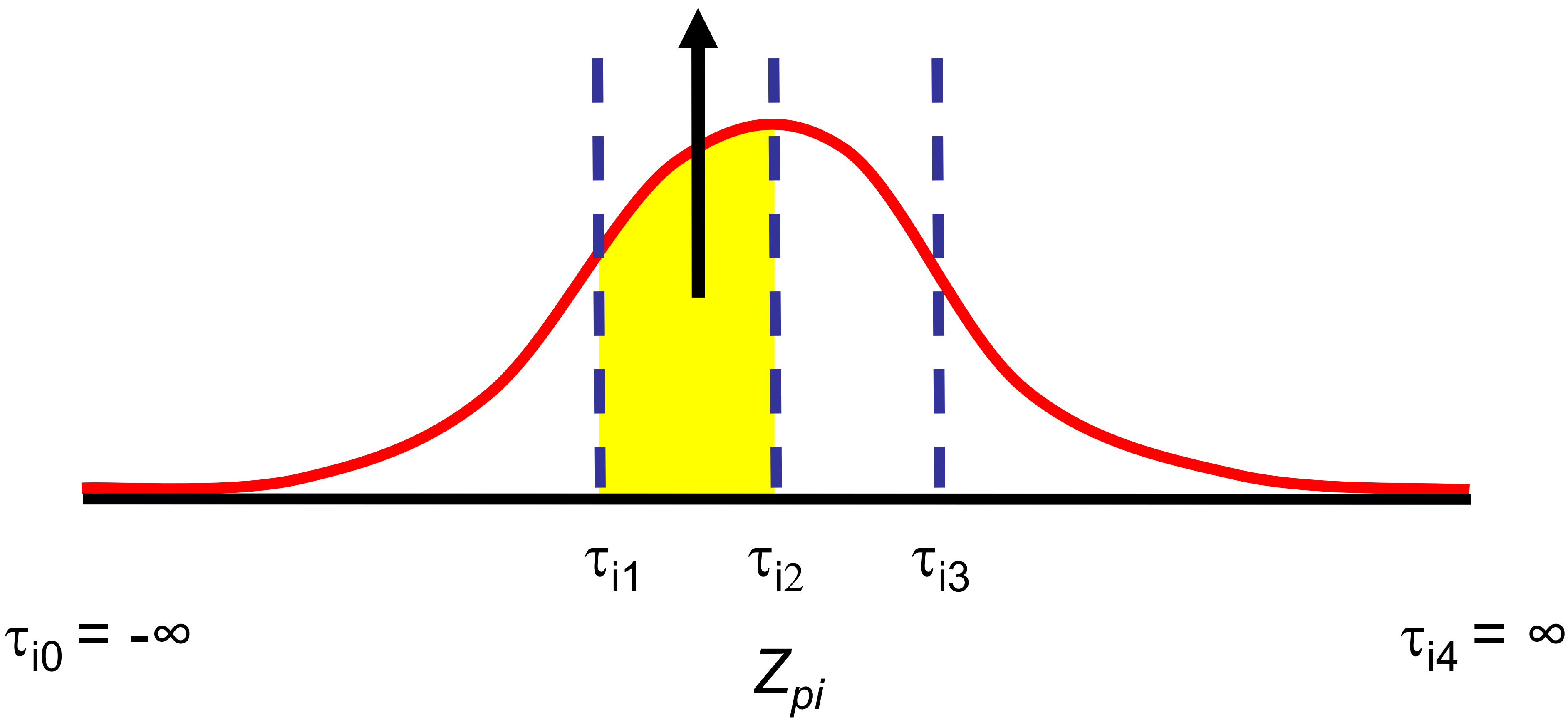
A Derivation of the GRM

(Wirth & E… …Leeuw, 1989)

- A continuous v… …s on item *i*.
- The ordinal sc… …holds.

Subjects score '1'

$$P(X_{pi} = 1 | Z_{pi}) = \int_{\tau_{i1}}^{\tau_{i2}} h(Z_{pi}; \boldsymbol{\gamma}_{pi}) dZ_{pi}$$



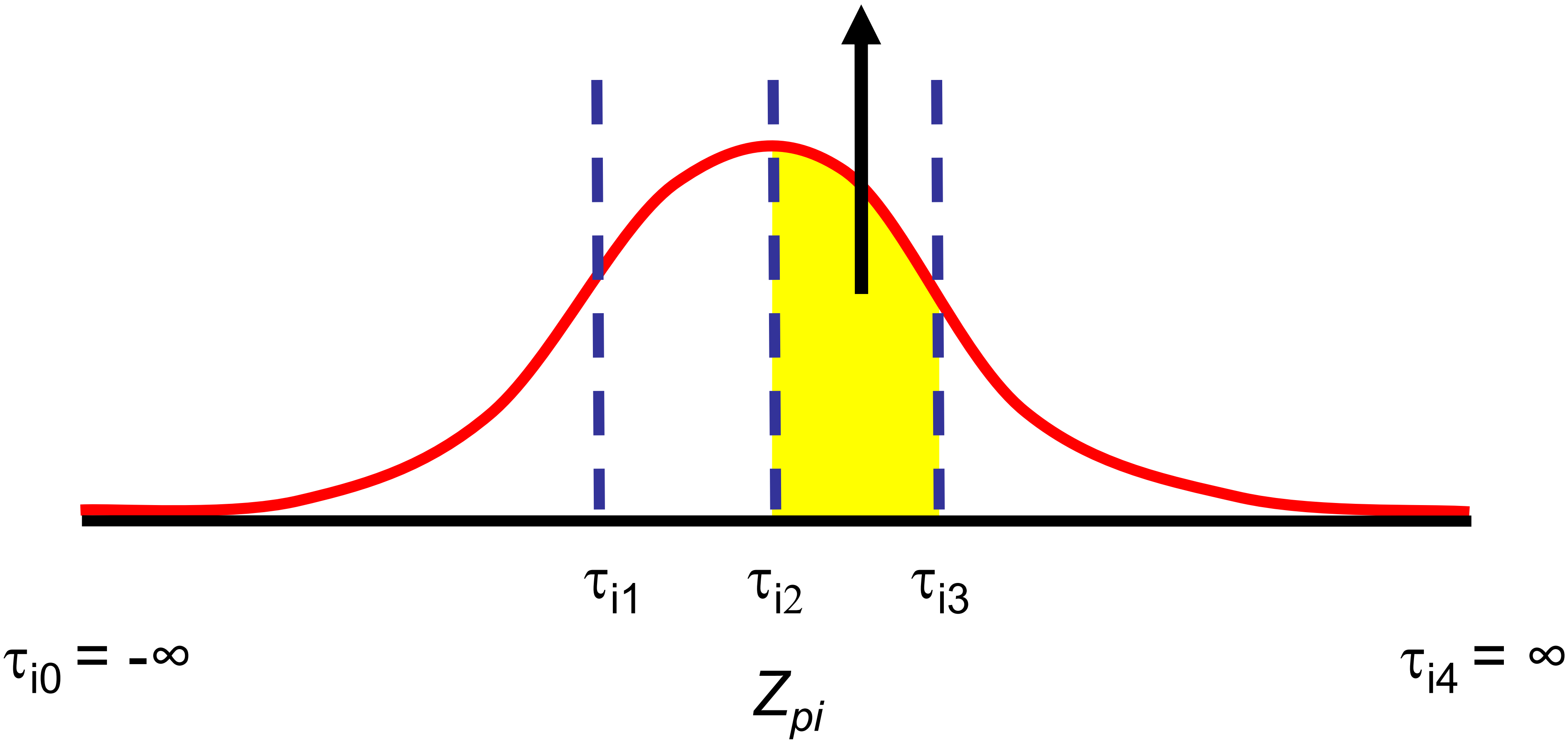
A Derivation of the GRM

(Wirth & Edwards, ... 1989)

- A continuous variable, Z_p ...
- The ordinal scores arise ...

Subjects score '2'

$$P(X_{pi} = 2 \mid Z_{pi}) = \int_{\tau_{i2}}^{\tau_{i3}} h(Z_{pi}; \boldsymbol{\gamma}_{pi}) dZ_{pi}$$



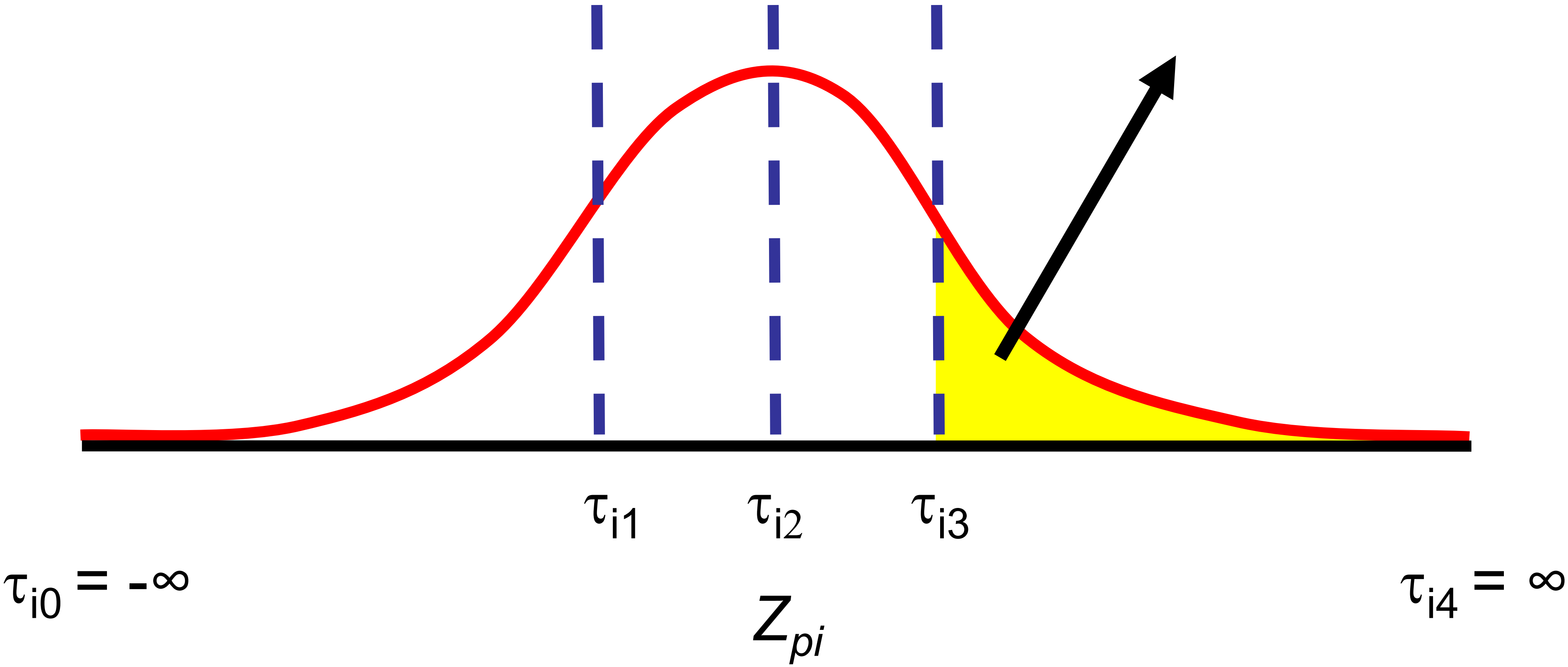
A Derivation of the GRM

(Wirth & Edwards, 2007; Takane & de Leeuw, 1989)

- A continuous variable, Z_{pi}, underlie
- The ordinal scores arise by cutting

Subjects score '3'

$$P(X_{pi} = 3 \mid Z_{pi}) = \int_{\tau_{i3}}^{\tau_{i4}} h(Z_{pi}; \boldsymbol{\gamma}_{pi}) dZ_{pi}$$



Generally:

$$P(X_{pi} = c \mid Z_{pi}) = \int_{\tau_{ic}}^{\tau_{i(c+1)}} h(Z_{pi}; \gamma_{pi}) dZ_{pi}$$

where *h*(.) is the density of Z_{pi} with parameter vector γ_{pi}.

Using a normal distribution for *h*(.): $Z_{pi} = \upsilon_i + \lambda_i \theta_p + \varepsilon_{pi}$

$$P(X_{pi} = c \mid \theta_p) = \Phi\left(\frac{\upsilon_i + \lambda_i \theta_p - \tau_{ic}}{\sigma_{\varepsilon i}}\right) - \Phi\left(\frac{\upsilon_i + \lambda_i \theta_p - \tau_{i(c+1)}}{\sigma_{\varepsilon i}}\right)$$

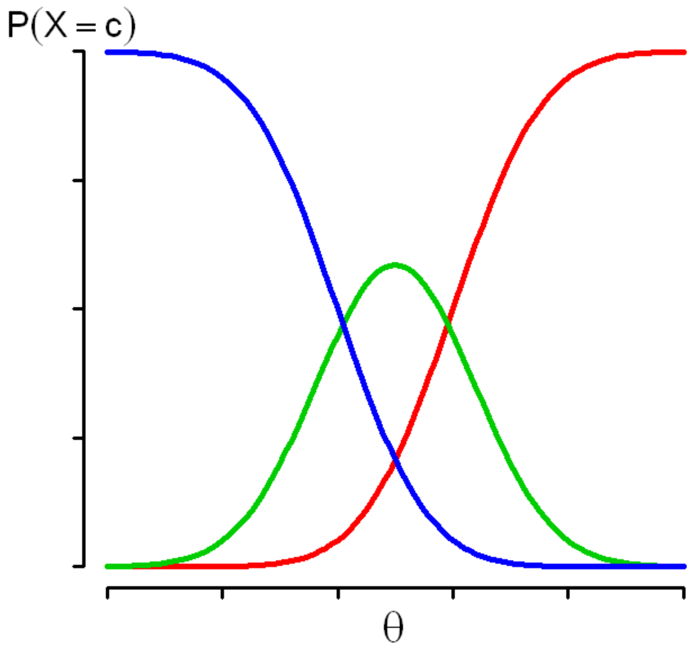
for: $\upsilon_i = 0, \quad \alpha_i = \frac{\lambda_i}{\sigma_{\varepsilon i}}, \text{ and } \beta_{ic} = -\frac{\tau_{ic}}{\sigma_{\varepsilon i}}$ (Takane & de Leeuw, 1987)

$$P(X_{pi} = c \mid \theta_p) = \Phi\left(\alpha_i \theta_p + \beta_{ic}\right) - \Phi\left(\alpha_i \theta_p + \beta_{i(c+1)}\right)$$

Distribution of Z_{pi}

- Traditionally, symmetrical distributions have been used (normal, logistic)

symmetrical



Asymmetrical distribution of Z_{pi}

- Asymmetry has implications for item information.
  - Implications for Computerized Adaptive Testing

- Samejima criticizes the symmetry of Z_{pi}
  - Inconsistent relation between β_{ic} and estimated θ_p, see Samejima (1997; 2000; 2008).

Tests on symmetry of Z_{pi}

- Samejima (1997; 2000; 2008):
  - Positive exponent family of GRM's
    - Acceleration parameter, ξ_c
    - Not applied yet.
- Bazan et al. (2006):
  - skewed Z_{pi} within 2 parameter normal ogive model
    - Found asymmetric Z_{pi} in a dataset on mathematical ability and weight perception.

Idea

If we accept the underlying linear relation between Z_{pi} and θ_p:

$$Z_{pi} = \upsilon_i + \lambda_i \theta_p + \varepsilon_{pi}$$

An asymmetrical distribution of Z_{pi} could be caused by

#1: Heteroscedastic residuals, $\varepsilon_{\mathrm{pi}}$.

#2: Asymmetrical distribution of θ_p.

We propose to test for these specific effects.

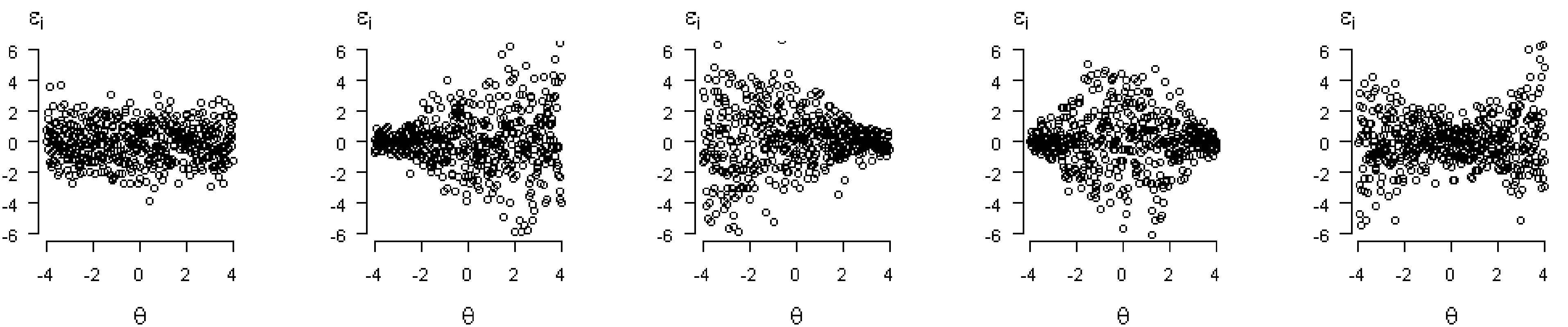
Why important

- As a test on the assumptions per se.
- Specific tests can be more powerful then marginal tests (Molenaar, Dolan, & Verhelst, 2010)
- Substantive applications
  - Ability differentiation (Spearman, 1927)
    - Stronger ability factor at lower levels of the ability
  - Moderation (Bauer & Hussong, 2009)
    - Item parameters are a function of an external variable, e.g., age

#1: Heteroscedastic residuals

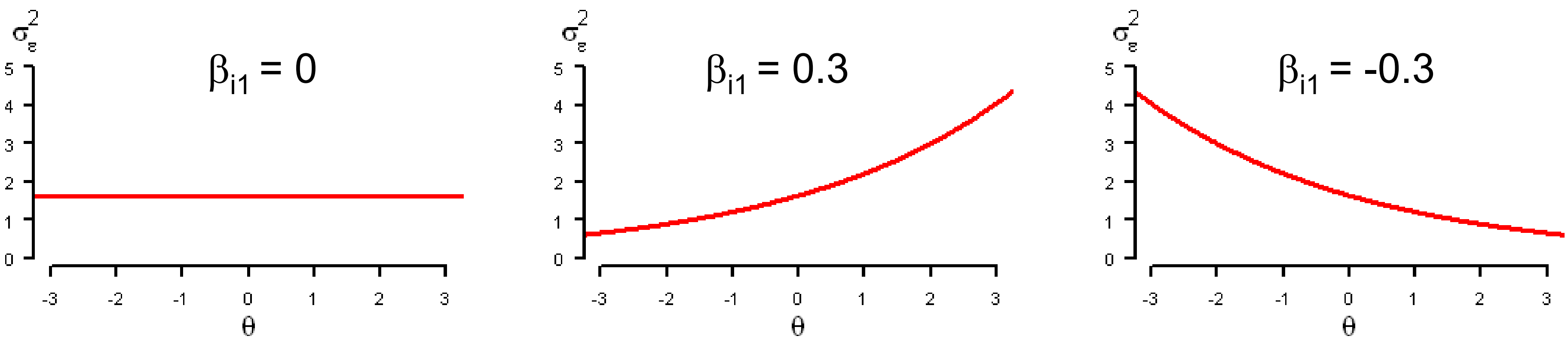
$$Z_{pi} = \upsilon_i + \lambda_i \theta_p + \varepsilon_{pi}$$

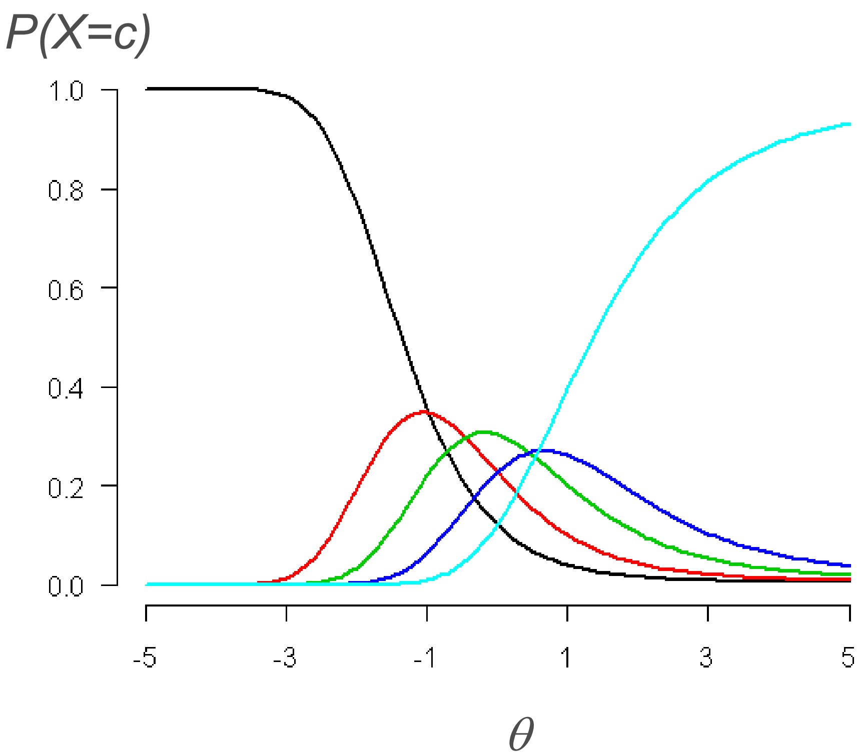
Make $\sigma_{\varepsilon i}^2$ a function of the latent trait, θ.



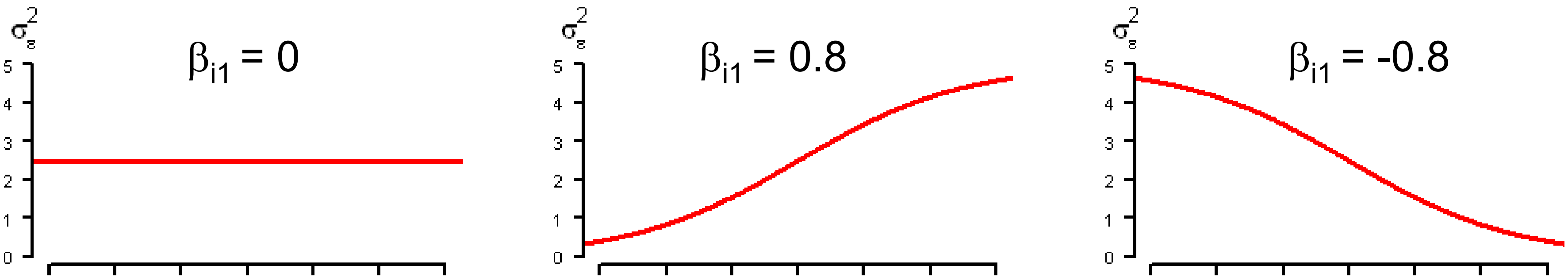
Heteroscedastic continuous factor model (Hessen & Dolan, 2009):

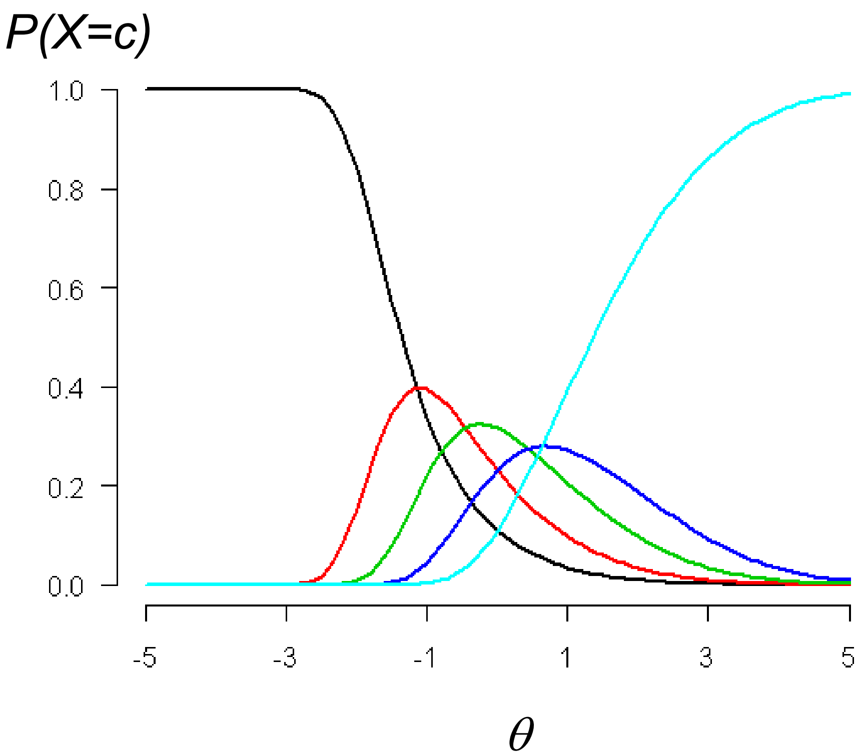
$$\sigma_{\varepsilon i}^2 = \exp(\beta_{i0} + \beta_{i1}\theta)$$





$$\mathrm{var}(\varepsilon_i) = \frac{2\beta_{i0}}{1 + \exp\left(-\beta_{i1}\dfrac{\theta - \mu_\theta}{\sigma_\theta}\right)}$$





#2: Skewed ability distribution

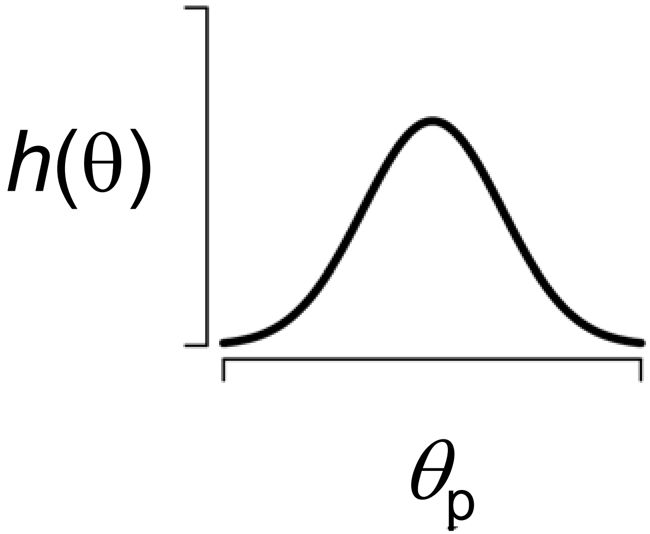
Skew-normal density (Azzalini, 1985; 1986)

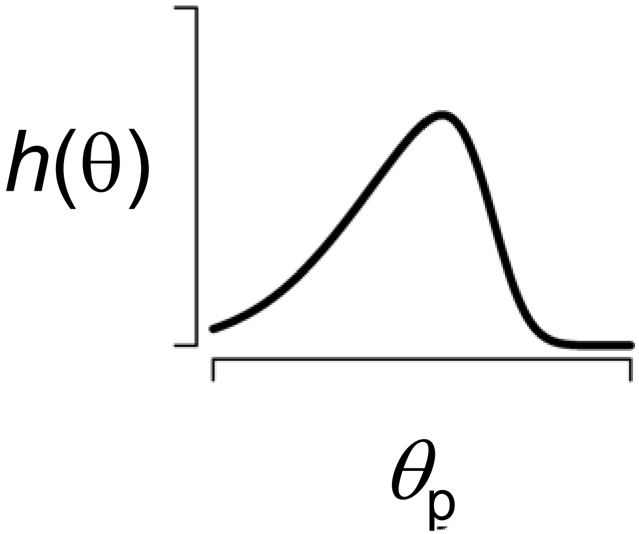
location parameter

$$h(\theta_p;\zeta,\omega,\kappa) = 2\Phi\left[\zeta\times\left(\frac{\theta_p-\kappa}{\omega}\right)\right]\times\frac{1}{\sqrt{2\pi}\omega}\exp\left[-\frac{1}{2}\left(\frac{\theta_p-\kappa}{\omega}\right)^2\right]$$

shape parameter *scale parameter*

Normal distribution function *Normal density function*



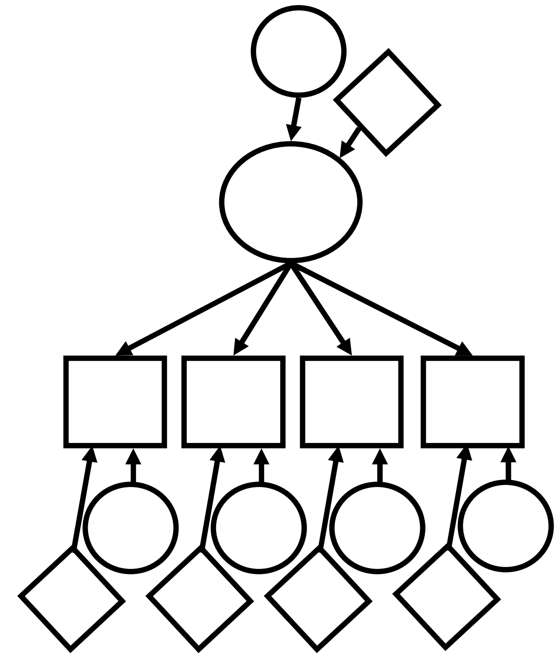


GRM model:

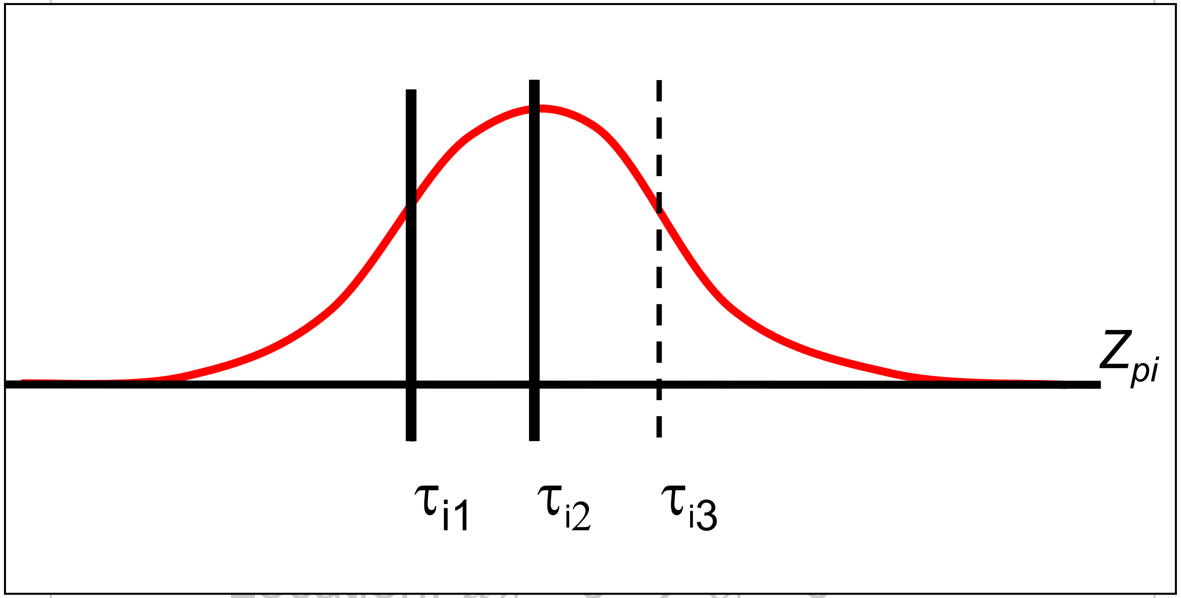
$Z_{pi} = \upsilon_i + \lambda_i \theta_p + \varepsilon_{pi}$

Traditional identification:

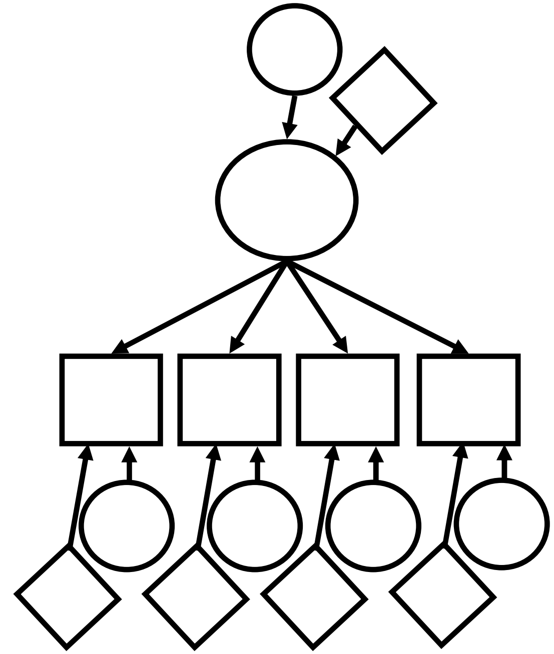
- Scaling θ
  - $\sigma_\theta^2 = 1$, $\mu_\theta = 0$
- Scaling Z_{pi}:
  - Location: $\mu_{Zpi} = 0 \rightarrow \upsilon_i = 0$
  - Scale: $\sigma_{Zpi}^2 = 1 \rightarrow \sigma_{\varepsilon i}^2 = 1 - \lambda_i^2$



Item factor model:



– Scale: $\sigma_{Zi}^2 = 1 \rightarrow \sigma_{\varepsilon i}^2 = 1 - \lambda_i^2$



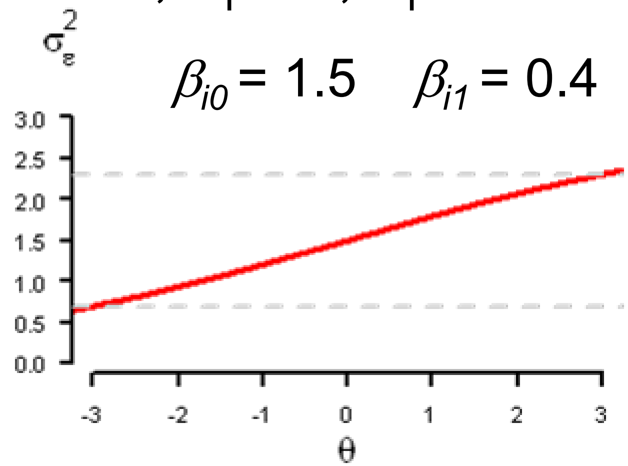
Alternative identification (Mehta, Neale & Flay, 2004):

- Scaling θ_p
  - $\sigma_\theta^2 = 1$, $\mu_\theta = 0$
- Scaling Z_{pi}:
  - **Fix two adjacent thresholds (e.g., τ_{i1} and τ_{i2})**

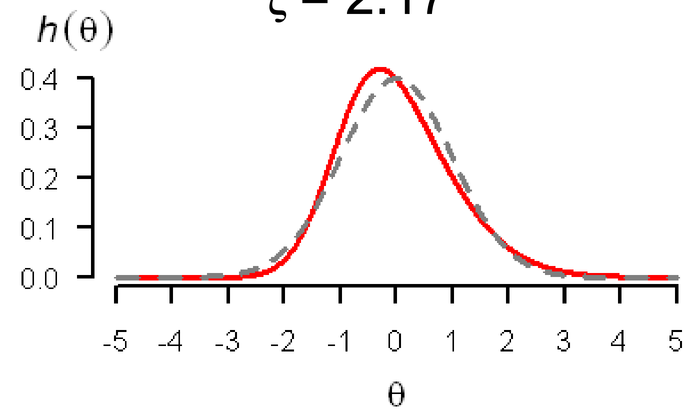
Simulation

- 10 items, $\upsilon_i = 0$, $\lambda_i = 1$

$\beta_{i0} = 1.5$ $\beta_{i1} = 0.4$



$\zeta = 2.17$



- Number of categories:

  3, $\tau_{i1}=-.75$ $\tau_{i2}=.75$

  5 $\tau_{i1}=-2,$ $\tau_{i2}=-.75,$ $\tau_{i3} = .75,$ $\tau_{i4} = 2$

- Number of subjects, 400 and 800.

- 50 replications, level of significance 0.05
- Marginal Maximum Likelihood (Bock & Aitkin, 1987) in the Mx software package (Neale, et al., 2007).

No subj	No cat	Effect	Hetero	Skew	Both
400	3	None	0.07	0.08	0.06
		Hetero	0.55	0.05	0.51
		Skew	0.14	0.4	0.29
		Both	0.26	0.38	0.29

Illustration

- Shortened PANAS (Positive Affect and Negative Affect Scale).
  - 557 psychology freshmen
  - 20 positive affect items, 20 negative affect items

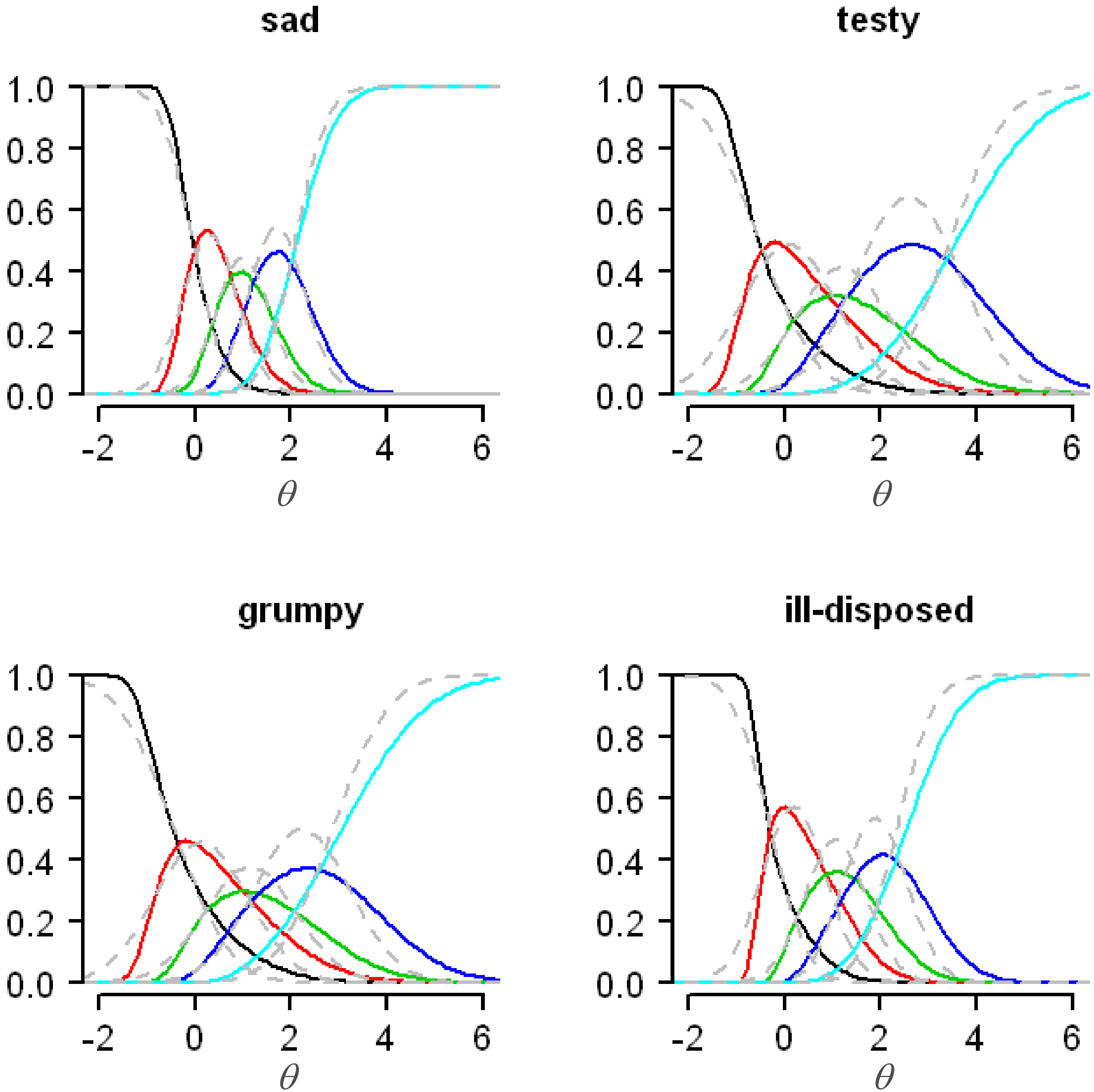
- Negative scale, e.g.,:

wistful	1	2	3	4	5
desperate	1	2	3	4	5
sad	1	2	3	4	5

- Traditional GRM, RMSEA: 0.066.

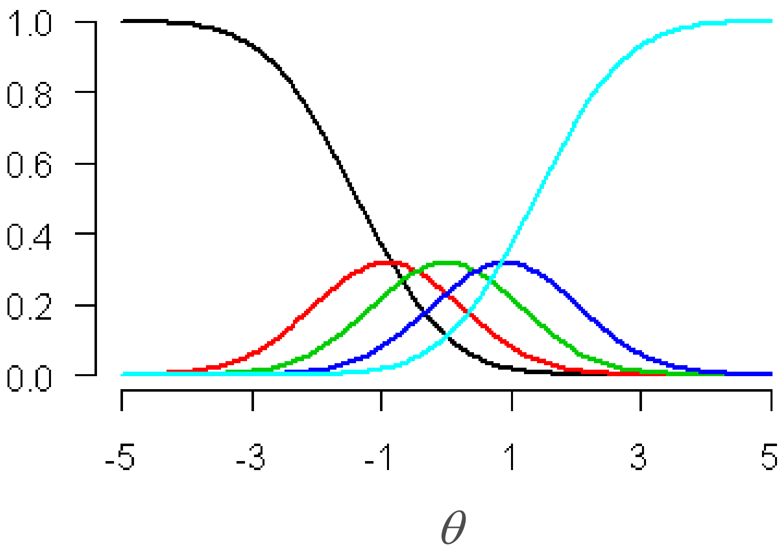
Illustration

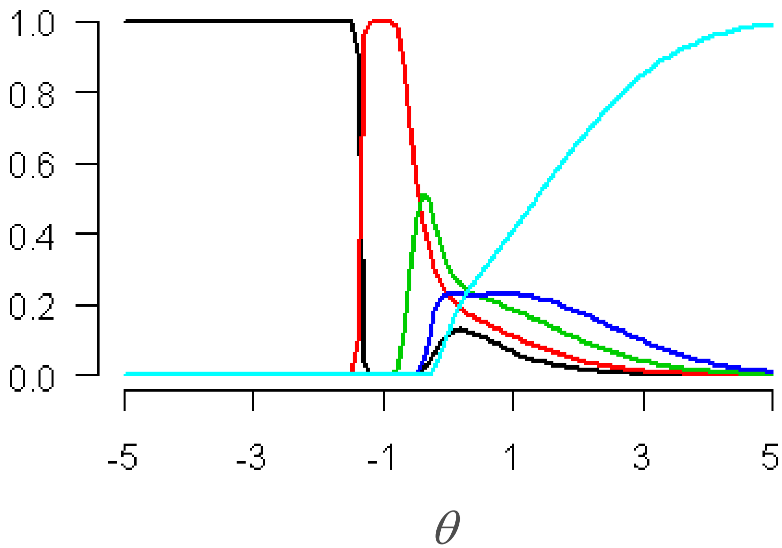
	LRT	AIC	BIC	sBIC	DIC
#1: full model	-	1716.31	-22701.99	-5399.54	-12684.64
#2: Skew dropped	$\chi^2(1) = 1.84$	**1716.14**	-22704.23	**-5400.20**	-12685.97
#3: hetero dropped	$\chi^2(20) = 87.97$	1764.28	-22721.23	-5387.04	-12685.50
#4: No effects	$\chi^2(21) = 87.99$	1762.29	**-22724.38**	-5388.61	**-12687.74**


sad
testy
grumpy
ill-disposed
1.0
0.8
0.6
0.4
0.2
0.0
-2
0
2
4
6
θ

Limitations

- Fixing two adjacent thresholds, you can make the residual variances a function of θ.
  - Exponential function misbehaves in the limit of θ.
  - Logistic function performs well in the limits of θ, but for very extreme effect sizes:





	Response Pattern	Normal Ogv.	Logistic
1	00000	neg. infinity	neg. infinity
2	10000	−2.28385	−2.28753
3	01000	−2.27016	−2.28753
4	00100	−1.84831	−2.28753
5	00010	−1.34811	−2.28753
[illegible]	01100	−1.15759	−0.75260
[illegible]	[illegible]	−0.86577	−2.28753
[illegible]	[illegible]	−0.75034	−0.75260
9	[illegible]	−0.75021	−0.75260
10	01010	−0.75013	−0.75260
11	00110	−0.75011	−0.75260
12	00101	−0.36062	−0.75260
13	10010	−0.34310	−0.75260
14	01001	−0.27309	−0.75260
15	00011	−0.19116	−0.75260
16	01110	−0.15292	0.75260
17	10001	0.15292	−0.75260
18	00111	0.19116	0.75260
19	01101	0.27309	0.75260
20	10110	0.34310	0.75260
21	01011	0.36062	0.75260
22	10011	0.75011	0.75260
23	10101	0.75013	0.75260
24	[illegible]	0.75021	0.75260
[illegible]	[illegible]	0.75034	0.75260
[illegible]	[illegible]	0.86577	2.28753
[illegible]	11001	1.15759	0.75260
28	10111	1.34811	2.28753
29	11011	1.84831	2.28753
30	11101	2.27016	2.28753
31	11110	2.28385	2.28753
32	11111	pos. infinity	pos. infinity

Answering the more **difficult** item correctly is given more credit

Answering the more **easy** item correctly is given more credit

	Response Pattern	$\hat{\theta}_v$
1	00000	neg.infinity
2	10000	-1.97543
3	01000	-1.72256
4	00100	-1.40982
5	00010	-1.29107
6	00001	-1.27506
7	11000	-0.44747
8	10100	-0.20737
9	01100	-0.18094
10	10010	0.09996
11	01010	0.11753
12	10001	0.21808
13	01001	0.23485
14	00110	0.24694
[illegible]	[illegible]0101	0.36499
[illegible]	[illegible]0011	0.67782
[illegible]	[illegible]1100	1.05493
18	[illegible]1010	1.29427
19	10110	1.32062
20	01110	1.32270
21	11001	1.60183
22	10101	1.61938
23	01101	1.62074
24	10011	1.74879
25	01011	1.74974
26	00111	1.76139
27	11110	2.56892
28	11101	2.81505
29	11011	2.84197
30	10111	2.84408
31	01111	2.84425
32	11111	pos.infinity

Answering the more **difficult** item correctly is given more credit